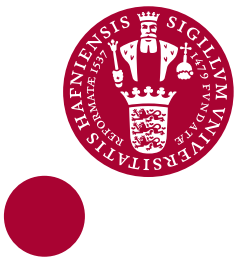


Structural basis of the bacteriophage TP9011 CI repressor dimerization and interaction with DNA

Rasmussen, Kim Krighaar; Varming, Anders Kokkenborg; Schmidt, Simon N.; Frandsen, Kristian Erik Høpfner; Thulstrup, Peter Waaben; Jensen, Malene Ringkjøbing; Lo Leggio, Leila

Published in:
F E B S Letters

DOI:
10.1002/1873-3468.13060

Publication date:
2018

Document version
Early version, also known as pre-print

Structural basis of the bacteriophage TP901-1 CI repressor dimerization and interaction with DNA

Kim K. Rasmussen, Anders K. Varming, Simon N. Schmidt, Kristian E. H. Frandsen, Peter W. Thulstrup, Malene Ringkjøbing Jensen, Leila Lo Leggio

FEBS LettersVolume 592, Issue 10, 1738-1750

which has been published in final form at https://doi.org/10.1002/1873-3468.13060**.**

Structural basis of the TP901-1 CI repressor dimerization and interaction with DNA

Kim Krighaar Rasmussen[1#], Anders K. Varming[1#], Simon N. Schmidt[1], Kristian E.H. Frandsen[1$], Peter W. Thulstrup[1], Malene Ringkjøbing Jensen[2], Leila Lo Leggio[1,*]

[1]Department of Chemistry, University of Copenhagen, Universitetsparken 5, Copenhagen, Denmark

[2]Univ. Grenoble Alpes, CNRS, CEA, IBS, F-38000 Grenoble, France

[$] Current address: INRA, UMR 1163 BBF (Biodiversité et Biotechnologie Fongiques), Polytech Marseille, Aix-Marseille Université, 163 Avenue de Luminy, 13288 Marseille Cedex 9, France

# equally contributing authors

[*] To whom correspondence should be addressed

Leila Lo Leggio

E-mail: leila@chem.ku.dk

Tel: +45 35 32 02 95

Running title: Characterization of CI-O_L repressor interaction

Keywords: phage repressor; protein-DNA complex; helical hook motif; dimerization domain; transcription factor

Abbreviations:

BCA: bicinchoninic acid

CAD: C-terminal anti-repressor-binding domain

CD: circular dichroism

CDD: Conserved Domain Database

CI: clear 1 repressor

CIΔ58: clear 1 repressor with 58 residues C-terminally truncated

CTD: C-terminal domain

CTD_1: C-terminal subdomain 1

DESY: Deutsches Elektronen-Synchrotron

ESRF: European Synchrotron Radiation Facility

HTH: helix-turn-helix domain

MOR: modulator of repression

MR: molecular replacement

NTD: N-terminal domain

O_D: operator site, distant

O_L: operator site, left

O_R: operator site, right

P_L: promoter, left directed

P_R: promoter, right directed

SAXS: Small angle X-ray scattering

SEC: size-exclusion chromatography

TP901-1: temperate phage 901-1

Abstract

Temperate bacteriophages are known for their bi-stability, which in TP901-1 is controlled by two proteins, CI and MOR. CI is hexameric and binds three palindromic operator sites via an N-terminal helix-turn-helix domain (NTD). A dimeric form, such as the truncated CIΔ58 investigated here, is necessary for high affinity binding to DNA. The crystal structure of the dimerization region (CTD_1) is determined here, showing that it forms a pair of helical hooks. This newly determined structure is used together with the known crystal structure of the CI-NTD and small angle X-ray scattering data, to determine the solution structure of CIΔ58 in complex with a palindromic operator site, showing that the two NTDs bind on opposing sides of the DNA helix.

Introduction

Bacteriophages are bacteria-feeding virus particles[1,2], and can be classified into two major classes, the lytic and the temperate groups. The group of lytic bacteriophages enters the lytic lifecycle directly after infecting a bacterium. Here new progeny is synthesized using the infected bacterium's transcription machinery. The synthesis of new progeny will eventually burst the infected bacterium releasing new progeny to the surroundings. Temperate bacteriophages are bi-stable, and thus have the ability to enter a second lifecycle, the lysogenic lifecycle, where the bacteriophage incorporates its genome into the genome of the infected bacterium, resulting in a prophage. The prophage will stay latent in the infected bacterial cell, and divide along with the bacterium[3].

A genetic switch is regulating the two lifecycles. One of the most studied genetic switches is that of the λ phage[4], but others have been studied as well[5,6]. In the temperate phage 901-1 (TP901-1) the genetic switch contains three palindromic operator sites (O_R; right, O_L; left and O_D; distant) and two divergently oriented promoters (P_R, right directed and P_L, left directed). P_R is responsible for the lysogenic lifecycle, and P_L is responsible for the lytic lifecycle. Furthermore, the genetic switch includes two genes (*mor*; modulator of repression and *cI*; clear I) coding for the two proteins, MOR and CI, which are the key players of the genetic switch.

The MOR protein contains 74 amino acids and it has been suggested to fold as a helix-turn-helix domain (HTH) [7]. A C-terminally truncated version of MOR has been shown *in vivo* to collapse the bi-stability of TP901-1, resulting in permanently repressed P_L and hence the lysogenic state[7]. Furthermore, it has been hypothesized that CI and MOR form a DNA-binding heteromeric complex with distinct specificity, thereby repressing the P_R promoter and resulting in the lytic lifecycle. The heteromer binding site on the DNA has yet to be identified, although theoretical studies suggest it to be downstream of the P_R promoter[8].

The other protein, CI, is composed of 180 amino acid residues and contains an N-terminal DNA binding domain (NTD, 1-80) with a HTH fold[9][10], and a C-terminal domain (CTD, 90-180) responsible for oligomerization[11]. A linker of nine residues (81-89) separates the two domains, and provides enough flexibility to ensure optimal DNA binding[9]. It has been suggested that CI binds as trimers of dimers to the three operator sites (O_R, O_L and O_D), in order to repress the P_L promoter[11]. C-terminal truncation studies show that the last 10 residues of CTD are necessary for formation of the hexameric arrangement, but the truncated variant retains the clonal variation ability[11]. The ability for clonal variation is first lost when CI is C-terminally truncated with 58 residues[11]. This truncated variant, CIΔ58 (1-122), forms a dimer consisting of NTD, linker and CTD_1 (90-122), a part of CTD recently suggested to be a subdomain of its own, and responsible for dimerization[9].

A search with the sequence of CIΔ58 using PHYRE2 [12], a fold-recognition program, resulted in

three top hits which also were DNA-binding proteins and dimerizing through a helical hook motif [9] [13]. However the sequence identity between the template PDBs and CTD_1 was extremely low, and alternative models (coiled coils of bent helices) were obtained using different bioinformatics approaches [9]. Here we present the crystal structure of TP901-1 CTD_1 finally establishing with certainty that it forms a dimer of helical hooks. Additionally, we structurally investigate the purified CIΔ58:O_L complex using small angle X-ray scattering (SAXS) and circular dichroism (CD) spectroscopy. Using the SAXS data and the crystal structures determined for the domains of CIΔ58, we construct a model for the solution structure of CIΔ58 in complex with a full O_L operator site, and show that CIΔ58 binds DNA canonically, with one NTD binding to each O_L half-site.

Materials and methods

Crystallization, data collection and structure determination of CTD_1

The CTD_1 peptide was purchased as a synthetic peptide from CASLO with a total of 35 residues. The purity of the peptide was 98.45% and it was delivered in a lyophilized chloride salt format without any modifications.

Former studies had revealed difficulties with solubilizing the peptide in a 20 mM NaF solution. This was overcome by dissolving the peptide in a buffer containing 20 mM Tris-HCl and 100 mM NaCl with a pH of 12. The pH was then lowered through buffer exchange using a 2 kDa cut-off Slide-A-Lyzer dialysis cassette (Pierce™) against the same buffer solution, but with a pH of 7.5. The protein concentration was measured the bicinchoninic acid (BCA) assay using kit from Pierce™ against a bovine serum albumin standard.

Sitting drop vapor diffusion crystallization screening was conducted with an Oryx 8 Protein Crystallization Robot and MRC 2 Well Crystallization plates (Douglas Instruments Ltd.) at room temperature using a JCSG+™ (Qiagen) and Morpheus® (Molecular Dimensions) screen. Each well consisted of a reservoir with a volume of 100 µL and a top and bottom drop, both with a volume of 0.3 µL but with protein to reservoir ratio of 3:1 and 1:1, respectively. All drops were set up with a stock protein solution of 4.4 mg/mL. Two conditions from the JCSG+™ screen were chosen for optimization: no. 40 (0.2 M lithium sulfate, 0.1 M sodium acetate, pH 4.5 and 30% w/v PEG 8000) and no. 45 (0.17 M ammonium sulfate, 25.5% w/v PEG 4000 and 15% v/v glycerol). Optimization was performed similarly as the screening but with drop volume of 0.4 µL (protein:reservoir ratio of 3:1 and 1:1, respectively, for top and bottom drop). The protein concentration was varied from 4.4 to 1.1 mg/mL and the PEG concentration was varied from 22-29% for condition 45 and 20-40% for condition 40 and the pH value was varied from 4-6 for condition 40. In general, condition 45 gave the best results with far more and higher scoring crystal hits. Finally the best crystals obtained from optimization were from same reservoir composition, as the original conditions.

All crystals were mounted on nylon loops (Hampton Research) with or without 22% (v/v) glycerol as cryoprotectant. Initial crystal tests were carried out at the ID30B and ID23-2 beamlines at the European Synchrotron Radiation Facility (ESRF, Grenoble, France), but final data were collected at the P11 beamline of Deutsches Elektronen-Synchrotron (DESY, Hamburg, Germany) at 100 K. The collected data sets were scaled and processed using the XDS package [14].

Two different crystal types were discovered when processing the datasets. The crystal grown from JCSG+™ condition 40 could be processed in the orthorhombic space group $P2_12_12_1$ as the highest symmetry, while crystals grown from condition 45 could be processed in the tetragonal space group $P4_12_12$.

Structure determination and analysis

Phasing information was obtained through molecular replacement (MR) using the AMPLE pipeline [15,16] after several failed attempts to use perfect α-helical templates as models. Given the very low level of sequence identities with the TP901-1 CTD_1, we did not extensively attempt to use the PHYRE2 models in MR searches. The *ab initio* models used as search models by AMPLE were produced using the QUARK server [17]. A successful solution was found by the AMPLE server using the tetragonal dataset. The statistics for the model showed a SHELXE correlation coefficient of 0.3799, R_{free} of 0.397 and an average chain length of 33 residues, all well beyond the threshold of a successful structure solution [15,16]. The AMPLE model consisted of two molecules creating a homodimer with all 35 residues modeled in the first monomer and the first 33 residues modeled in the second. However, further improvement of this model proved difficult, as did further structure refinement attempts with other datasets, which had been processed in the tetragonal space group.

Instead, the output model from AMPLE was used as search model for molecular replacement in Phenix.Phaser[18] with the orthorhombic dataset searching for three dimers in the asymmetric unit. The resulting output model was chosen for further refinement.

Automated model building and computational refinement was carried out using Phenix.Autobuild[19] and Phenix.refine, respectively, while manual refinement and structure building were performed using Coot [20,21]. Validation was performed using MolProbity[22] and structure analysis using PDBsum and PISA[23,24]. Search for structural homologues was carried out using DALI[25].

Preparation of ds-O_L for SAXS and CD studies

Anti-sense (TCAAGTACTTTGCACTTGA, 5778 Da) and sense (AGTTCATGAAACGTGAACT, 5836 Da) were purchased from TAGCopenhagen. Each ss-O_L 19'mer was solubilized in annealing buffer (10 mM

Tris-HCl, 50 mM NaCl, 1 mM EDTA, pH 7.5) in a sufficient volume to obtain 2 mM stocks. Equal volumes (1:1) were mixed to obtain a 1 mM ds-O_L (11614 Da). The sample was incubated at 368.15 K for five minutes and the heat block containing the sample was allowed to cool to room temperature on the workbench.

Expression and purification of CIΔ58 and CIΔ58:O_L complex

The expression and purification of CIΔ58 a truncated version of CI, containing 122 residues of CI from TP901-1, with two additional residues from cloning and a His-tag$_6$ at the C-terminus (RSHHHHHH, 30.088 kDa), have been described previously [9,10,26]. 980 µL purified CIΔ58 protein at 44.54 µM (monomeric CIΔ58 M_w = 15.05 kDa) in (20 mM Tris, 100 mM NaCl, pH 7.5) was mixed in a ratio CIΔ58 dimer to O_L 1:1.2 ,with O_L (in annealing buffer, 20 µL of 1 mM dsO_L) and incubated for an hour at room temperature. The CIΔ58:O_L complex was purified by size-exclusion chromatography (SEC) using a HiLoad 26/60 Superdex75 with elution buffer (20 mM Tris, 100 mM NaCl, pH 7.5). M_w was estimated from the standard curve previously published [9] as shown in the Results section and associated figures.

Samples were concentrated using Amicon Ultra Centrifugal filters with a molecular weight cut-off of 10 kDa. The CIΔ58 concentration was estimated from the UV absorbance at 280 nm using a theoretical extinction coefficient of 8480 $M^{-1}cm^{-1}$ calculated using the ExPASy tool ProtParam [27]. Concentration of CIΔ58:O_L was either determined by measuring O_L concentration with A_{260} nm (neglecting the protein contribution), assuming that a dimeric CIΔ58 and dsO_L interact 1:1 or by determining the CIΔ58 concentration with BCA measurements. The two methods deviated by 2.3 % indicating that the assumption was correct, and the CIΔ58:O_L interact 1:1 as also shown previously [9].

DNA retardation gels

To verify if the desired protein:DNA complex had been purified we used 6 % Novex® tris base, boric acid, EDTA (TBE) gels (ThermoFisher) using 0.5 X TB as running buffer. A total volume of 10 µL of each sample was loaded with suitable sample concentration. The current was adjusted to between 12-15 mA at the start, and run for 30 min in cold room (+4 °C). Hereafter, the gel was stained for 10 min. in ethidium bromide, and afterwards stained in InstantBlue® overnight.

SAXS data collection, processing and modelling

SAXS data of CIΔ58, O_L and the CIΔ58:O_L complex were collected at station BM29 at the ESRF Grenoble, France using a PILATUS detector. Data were collected using a sample to detector

distance of 2.87 m and a temperature of 277.15 K covering a range of momentum transfer of $0.05 < s < 5$ nm^{-1}, where $s = 4\pi\sin\theta/\lambda$, 2θ is the scattering angle and $\lambda = 0.992$ Å. CIΔ58, O_L and CIΔ58:O_L complex data were collected at BM29. An absolute calibration was carried out using water. While the CIΔ58 concentration was determined as before[9], concentration of O_L and CIΔ58:O_L complex were estimated based on A_{260} nm and BCA: CIΔ58; 0.69 mg/mL (0.02293 mM), 1.34 mg/mL (0.04454 mM), O_L; 0.163 mg/mL (0.01396 mM), 0.307 mg/mL (0.026 mM), 0.556 mg/mL (0.048 mM), 1.225 mg/mL (0.105 mM)) and CIΔ58:O_L were (0.333 mg/mL (0.008 mM), 0.466 mg/mL (0.0112 mM), 0.577 mg/mL (0.0138 mM), 0.88 mg/mL (0.0211 mM)).

SAXS data were background subtracted using the data processing program PRIMUS [28]. From the background scattering curve, the forward scatter, $I(0)$, and the radius of gyration, R_g, were obtained either with the Guinier approximation [29] or with the entire scattering profile using AUTOGNOM [30]. The molecular mass, M_w, was estimated both from the Guinier approximation and by calculating the excluded volume, V_p, of the hydrated particle from the Porod equation[31] in AUTOPOROD[30]. The pair distribution function of the particle, $P(r)$, and the maximum particle dimension, D_{max} were obtained from AUTOGNOM.

Twenty *ab initio* models of CIΔ58 were reconstructed using GASBOR22i[32] with default settings except when P2 symmetry was imposed. The reconstructed models were aligned with SUPCOMB20[33] and an average by DAMAVER[34]. DAMFILT[34] was used to represent the average *ab initio* model as a more compact and probable model. Twenty reconstructed *ab initio* models of CIΔ58:O_L complex were calculated using MONSA[35] with default settings except when P2 symmetry was imposed. From the twenty models, models with similar trends were selected. The unselected models fit the experimental data worse, and did not make sense chemically *e.g.* the DNA was attached to in the CTD1 region, or the CIΔ58 was separated into two monomers. An average *ab initio* model was calculated as described above.

CD spectroscopy

To study the interaction of CIΔ58 with O_L DNA, purified CIΔ58 was transferred to a buffer suitable for both gel filtration and CD spectroscopy (10 mM Tris base, 100 mM NaF, pH 7.5). The CD experiments were conducted on a Jasco J-815 CD spectropolarimeter in a suprasil quartz cell with 1 mm light path (Hellma Analytics). The spectra were run at room temperature in continuous mode at 20 nm/min from 305 nm to 180 nm, averaging five spectra per measurement. Care was taken to obtain and apply correction for references containing buffer and/or O_L DNA.

Results and discussion

The crystal structure of CI-CTD$_1$, the dimerization region of CIΔ58, reveals a helical hooks fold

In this study, the truncated dimeric form of CI, CIΔ58, serves as a model for the interaction of CI and its operator sites. This variant has recently been structurally characterized by SAXS, NMR, and CD spectroscopy in its unbound form [9]. The crystal structure of NTD in complex with an operator half-site is known[10] , while the structure of CTD has so far eluded. Here we present the structure of the first subdomain of CTD, CTD$_1$, responsible for dimerization.

Two crystal types of CTD$_1$ were produced – one tetragonal and one orthorhombic – which diffracted to 1.8 and 1.5 Å, respectively. The tetragonal form was used in MR with *ab initio* modeling, producing a dimeric structure as solution. Although the structure could not be refined to acceptable R-factor and R-free, this dimeric structure could be used as a search model in MR for the higher-diffracting orthorhombic form.

In both crystal forms the CTD$_1$ dimer resembles a pair of helical hooks (Fig. 1). In the orthorhombic crystals an imperfect trimer of dimers constitutes the asymmetric unit (Fig. S1). In the tetragonal crystal form, which led to the initial structure determination with AMPLE, there is a dimer in the asymmetric unit. However, refinement of this crystal form gave at best R_{work}/R_{free} of 0.294/0.345 and no further improvement could be obtained. Crystallographic symmetry generates an additional dimer, which is positioned as in the orthorhombic asymmetric unit. However, there is no convincing density for a third dimer, and attempts to place an additional monomer in the tetragonal asymmetric unit by MR failed. Placing an additional monomer manually in the tetragonal asymmetric unit using the orthorhombic positioning as a guide and then applying crystallographic symmetry, shows that a dimer can almost be formed, but is imperfect. Thus, it is possible that additional dimers are present in the tetragonal crystals but break the symmetry, and thus are disordered. Refinement of the tetragonal form was therefore abandoned.

The orthorhombic structure could be fully refined with anisotropic B-factors for all non-H atoms. The final structure had R_{work}/R_{free} of 0.145/0.197 (see additional statistics in Table 1). The quality of the electron-density map was excellent except for some flexible side chains and terminal residues. All residues could be modeled in all six chains except for residue 35 in chain D. The six chains have different numbers of side chains in alternate conformations (1 to 5) and the all atoms/Cα RMSDs between chain A and the following five chains were 0.949/0.339, 0.793/0.321, 0.958/0.408, 1.285/0.531 and 1.476/0.209 Å, respectively. The final model included 219 water molecules and three sulphate ions.

Each chain forms two helices comprising 11 and 20 residues, respectively, connected by a turn of two residues (Fig. 1). The dimers are held together by a slightly varying number of interactions. The dimer

formed by chain A and B has 25 residues from each chain interacting with an interface area of 1,325 $Å^2$ and 154 non-bonded contacts, 15 hydrogen bonds and two salt bridges (Fig. S2). The interaction is thus mostly hydrophobic with most of the interacting residues in the core of the helices being aliphatic or non-polar and the residues in the termini of the helices being polar and/or charged. The 15 hydrogen bonds are primarily formed between residues in the termini of the second helix of each chain and two salt bridges are formed between residues R106 of chain A and E120 of chain B and vice versa (Fig. 1), which suggests that the second helices in each monomer dictate the specificity of the dimer interaction.

PISA[23] analysis suggests that both the trimer of dimers and the dimer could be relevant biological units. We know however that the CIΔ58 construct does not form hexamers[9,26]. Furthermore, the trimer (Fig. S1) is not symmetric, as the interactions are not the same between all dimer pairs, and it is not stable enough to clearly observe in the tetragonal crystal form. Therefore, while the biological relevance of the dimer is supported by other biochemical and biophysical studies, the relevance of the hexameric assembly of CTD_1 alone in the crystal is not.

Relation of CTD_1 to other helical hooks

Proteins with a similar motif in the C-terminal region were previously found by conducting a PHYRE2 search targeting the primary sequence of CIΔ58[9]. The three top hits were the transcriptional regulator from *Listeria innocua* (PDB entry 3OP9, *Li*TR), the possible transposon-related DNA-binding protein from *Clostridium difficile* 630 (PDB entry 3IVP, *Cd*TR) and the restriction-modification controller protein C.Csp231I (PDB entry 3LIS, *Cc*RM). Superposition of the monomers from these three structures with a monomer of CTD_1 gave Cα RMSDs of 0.944, 1.328 and 2.060 Å, while superposition of the dimers gave Cα RMSDs of 1.436, 1.897 and 3.857 Å for 3IVP, 3LIS and 3OP9, respectively. When structurally aligning the monomers, differences are mostly due to the angles between the two helices. When structurally aligning the dimers, the larger differences are due to the relative orientation between the monomers. The structural similarity is however still remarkably high considering the low sequence similarity. The 5^{th} and 6^{th} ranking hits in this PHYRE2 search, PDB codes 1B0N and 3QYX, also contained helical hooks dimerization regions. 1B0N is a heterodimer of SinR, the master regulator of biofilm formation in *Bacillus subtilis,* with SinI [36], but similar dimers (in this case homodimers) of SinR [37] are found in PDB entry 2YAL, as part of a homotetramer. 3QYX is a similar dimerization domain associated with a HTH motif DNA binding domain found in EspR, the ESX-1 secretion system regulator found in *Mycobacterium tuberculosis* [38]. For these

and other helical hook proteins mentioned from now on, structural comparison parameters with CTD_1 are given in Table S1. For selected examples, the helical hooks structures are shown in Fig. S4.

A protein structure comparison was additionally conducted using the DALI server[25]. The top hit with a Z-score of 4.6 was the possible transposon-related DNA-binding protein from *Clostridium difficile* 630 (PDB entry 3IVP), also found from the PHYRE2 search. A superposition with CI-CTD_1 is shown in Fig. S4.

The following hits with a Z-score over 4.0 are all NblA proteins involved in degradation processes in cyanobacteria and red algae (PDB entries 1OJH, 2QDO and 2Q8V). They consist of a helix-loop-helix, which dimerizes into a four-helical bundle, but this motif is more open and has a completely different orientation compared to the CTD_1 dimer (see the motif in 2Q8V as example, in Fig. S4).

One hit with a Z-score of 3.8 is the N-terminal dimerization domain of VicH, a histone-like nucleotide structuring modulator of gene expression from *Vibrio cholera* (PDB entry 1OV9). The dimerization domain is helical and resembles the CTD_1 dimer to some extent, but one of the helices in each monomer is bent in an almost perpendicular angle (see Fig. S4). The next hit with a Z-score of 3.7 is a flagellar protein with unknown function found in bacteria (PDB entry 1VH6). It forms a dimer but each monomer consists of four helices, and the dimer interface is not similar to that of the CTD_1 dimer. None of the remaining 100 hits with a Z-score above 3.0 contained a dimerization domain resembling that of CTD_1.

We sought to identify additional helical hook dimers of transcription factors and regulators by a new PHYRE2 search and visual inspection of members of the Cd00093 HTH_XRE family in the Conserved Domain Database (CDD) [39], into which the NTD of TP901-1, *Li*TR, *Cd*TR and *Cc*RM are classified: BswR[40], a dimeric controller of motility and biofilm formation in *Pseudomonas aeruginosa* (PDB entry: 4O8B), and a repressor protein from *Salmonella* temperate phage (PDB entry: 5D4Z) could be additionally found[41]. The latter can be considered a helical hook dimer containing a short (170-175) and a long (178-196) helix connected by a short loop. The angle between the two helices and the orientation between the two monomers differ considerably from those of the CTD_1 dimer. Moreover, the dimer interface consists of only 40 non-bonded contacts and neither salt bridges nor hydrogen bonds [42].

Except for the longer *Salmonella* phage repressor, all these proteins containing helical hook dimers, *Li*TR, *Cd*TR, *Cc*RM, SinR, SinR/SinI, EspR, VicH, and BswR contain a maximum of 132 residues in contrast to the CI repressor from TP901-1, which contains 180 residues, and they do not have additional domains after the dimerization region. Length of the helices, the angles between them, and the relative orientation of the two pairs of intertwined α-helices, length and flexibility of the linker diverge considerably, as do the overall arrangement of the dimers, which in turn could influence the potential for binding DNA.

For all similar helical hooks (3IVP, 3OP9, 3LIS, 1B0N, 3QYX, 2YAL, 1OV9, 4O8B and 5D4Z) the dimer interface was analysed as for CTD_1 (Table S1). This analysis highlights that despite using a similar structural motif, there are many differences in the details, and there is no general structural strategy for the stabilization of the dimer. Even for the closest structural homologue *Cd*TR, the polar interactions and interacting residues are not particularly conserved (Fig. S3).

Purification and SEC of the CIΔ58:O_L complex

Following immobilized metal ion affinity chromatography purification, CIΔ58 was further purified using SEC (Fig 2, A). The protein eluted at 175.97 mL, and corresponds to a dimer of CIΔ58 with a typical A_{260}/A_{280} ratio for pure protein (Fig 2, C). The A_{260}/A_{280} ratio can be used to estimate the purity of a sample. An A_{260}/A_{280} = 0.57 corresponds to pure protein, whereas A_{260}/A_{280} = 2.0 corresponds to pure DNA. A complex is expected to be between A_{260}/A_{280} = 1.4-1.8. The purified CIΔ58 dimer (30.10 kDa) was mixed with dsO_L (11.61 kDa) and incubated for 60 min. at room temperature. The mixture was purified on same SEC column (Fig 2, B). Two major peaks were observed: Peak 1 (156.57 mL) and Peak 2 (180.85 mL). Using the previously published standard curve[9], we estimated the M_w of Peak 1 and $Peak_2$ to 48.31 kDa and 30.85 kDa, respectively, suggesting that Peak 1 is the CIΔ58:O_L complex, though there is a deviation between observed and estimated M_w (41.72 kDa) of 14 %. Peak 2 could correspond to CIΔ58, however the higher A_{260}, DNA retardation gels (see below) and BCA analysis suggests that Peak 2 only contains DNA. Thus, we did not focus on Peak 2.

DNA retardation gel

Samples from SEC purification Peak 2 and Peak 1 (Fig 3. lanes 6 and 7, 8, 9, respectively) were, together with two controls (pure O_L and CIΔ58 Fig 3. lanes 1, 2 and 4, respectively), loaded onto a DNA retardation gel and stained with EtBr (Fig. 3 top) and InstantBlue® (Fig. 3 bottom) to detect if the fractions contained both O_L and CIΔ58. Of the two controls, only O_L enters the gel (lane 1 and 2). The pI of CIΔ58 is probably the reason for not observing CIΔ58 as it is not capable of entering the gel alone (lane 4). From the gel it is evident that Peak 2 (although very weak) contains free O_L in good agreement with BCA measurements, which detected no protein. From the gel it is furthermore evident that Peak 1 contains both O_L and CIΔ58, and that they are in complex as CIΔ58 has entered the gel due to different charge of the CIΔ58:O_L complex compared to free CIΔ58 (lane 7, 8 and 9). This is also in good agreement with BCA measurements performed to determine the CIΔ58:O_L concentration.

SAXS studies of the CIΔ58:O_L complex

We have previously attempted to crystallize the CIΔ58:O_L complex, but without success, and as shown recently the flexibility of the complex may hinder crystallization[9]. Thus, solution studies (SAXS) were performed to obtain structural information of the truncated variant of CI in complex with the O_L site of the genetic switch from TP901-1. As the complex was purified at 4 °C, SAXS measurements were also carried out at 4 °C. The scattering profiles (Fig. 4 A) of the three scattering particles (CIΔ58:O_L, CIΔ58 and O_L) are, as expected, different. Clear differences are also seen in the Kratky plots (Fig. S5 A). Here the unbound CIΔ58 clearly has two maxima, when compared to the CIΔ58:O_L. In both cases however, flexibility is indicated by the increasing values of I $*s^2$ at high s (> 3). In contrast the O_L is a rigid scattering particle, with no flexibility, as seen from the decreasing values of I $*s^2$ at high s. From the Guinier region (Fig. S5 B) it was found that none of the samples (CIΔ58:O_L, CIΔ58 and O_L) showed attractive, repulsion or aggregation tendency in the measured concentration range. Thus, highest concentration for each scattering particle was used for analysis and modelling. R_gs for all three scattering particles (CIΔ58:O_L, CIΔ58 and O_L) were obtained by the Guinier approximation (2.88, 2.82 and 1.91 nm), and from AutoGNOM [43] (2.93, 2.93, 1.98 nm) (Table 2). Note that for CIΔ58, the radius of gyration (R_g) is slightly smaller than previously reported at room temperature and 1 M NaCl (3.1 nm), but the P(r) functions (Fig. S6) have similar features. At 4 °C and 0.1 M NaCl, the R_gs indicates that CIΔ58 alone is more extended than CIΔ58 in complex with O_L as they have similar R_gs, although a M_w difference of ~33 %.

The P(r) curve of the CIΔ58:O_L has a single maximum at ~2.49 nm. Compared with the P(r) of free CIΔ58, which has two maxima, a global one at 2.14 nm and a local one at ~4-5 nm (Fig. 4 B), this indicates a structural change upon complex formation due to the presence of O_L. Structural changes are also observed from the Kratky plot.

The arrangement of the CIΔ58:O_L complex was determined by reconstruction of *ab initio* models from the P(r) curve (Fig. 4 C-E). As seen from the reconstructed MONSA *ab initio* model of CIΔ58:O_L the cross section (thickness of models) is comparable with the cross section of the free CIΔ58 (2.14 nm). The distances between the two NTD domains deviated (as expected) between bound and unbound CIΔ58 protein. The NTDs center of masses in the flexible unbound CIΔ58 are on average 5.5 nm from each other as previously found[9], whereas the distance between center of masses between NTDs in the bound CIΔ58 is ~ 3.0 nm, in agreement with the distance between two major grooves in B-DNA (3.4 nm). The reconstructed *ab initio* model shows that the NTDs bind to DNA from opposite sides, as expected for binding to a palindromic sequence. Thus we can establish with certainty that TP901-1 CIΔ58 binds canonically to DNA.

One of the major differences between bound and unbound CIΔ58 is that in the bound CIΔ58 the central region corresponding to the dimerization subdomains is elongated. This could indicate either a more rigid His-tag or structural changes in the helical hook dimer region, for example to a more coiled-coil structure. To determine if a coiled coil conformation described experimental data better than the helical-hook conformation found in the crystal structure, we prepared two atomic models, one with the dimerization region as observed in the crystal structure and one modelled as coiled coil dimerization, and fitted the theoretical scattering curves to the experimental data (Fig. S7). The model with helical hooks dimerization region describes experimental data significantly better than coiled coil, χ^2 =1.78 and χ^2 = 8.22, respectively. At this point therefore the helical hook dimer remains the best available model for this region also in the complex.

CD spectroscopy studies of CIΔ58:O_L interactions

CD spectroscopy was applied to observe the DNA interaction and the possible associated structural changes in solution. The changes in CD spectra were recorded for purified CIΔ58 with and without the 19mer O_L dsDNA oligonucleotide. CD spectra were measured for DNA alone (without protein), see Fig. S8. The final concentrations at the end of the titration were 1.9 μM CIΔ58 (monomer):1.9 μM (ds O_L) (excess DNA). Figure 5 shows the reference-corrected CD spectra for the mixture of protein and DNA, the mathematical sum of the protein and DNA spectra taken individually, and finally the difference between the experimental and summed data. The clear difference is indicative of a CD-sensitive structural change on formation of the protein-DNA complex. The minor offset from zero in the near-UV region shows that there is a slight stabilization of DNA on complexation. In the far-UV the spectra are dominated by the electronic transitions of the protein backbone, and here the negative contribution between 240 nm – 200 nm and the positive contribution below 200 nm point to an increased alpha-helicity of the protein in the DNA complex. The secondary structural content of CIΔ58 was assessed by spectral convolution using Dichroweb [44]. The results showed a slightly increased amount in α-helical content in CIΔ58 interacting with DNA (Table S2).

Concluding remarks

We have now experimentally determined the high-resolution crystal structure of the dimerization subdomain of TP901-1 CI, CTD_1, and shown that it folds as a pair of helical hooks. Through different approaches, we have found similar helical hook dimers in a number of other regulatory proteins with HTH DNA-binding domains, mostly from bacterial regulatory systems, but also one phage repressor. While the search methods may be slightly biased towards regulatory proteins, and our definition of structural similarity somewhat

restrictive, it does seem that this very simple motif is highly represented in DNA-binding regulatory bacterial proteins, but hardly present in, for example, enzymes. The role of the helical hook is clearly to favor dimerization, sometimes in the context of a larger oligomeric assembly. In at least one system, the SinR/SinI system regulating biofilm formation in *B. subtilis*, the helical hook interaction is also directly involved in the repressor/antirepressor interaction and not solely dimerization of the repressor[37].

The overall arrangement of these helical hooks and their dimer interfaces vary considerably, which makes it difficult to identify the motif by simple sequence identity. The sequence diversity makes it difficult to ascertain whether the structural motif has spread through the bacterial and phage world by divergent rather than convergent evolution, yet the prominence in certain type of regulatory systems of bacteria and phages seems in favor of the former. The helical hook region is often associated with and separated by a variable linker from an N-terminal HTH DNA binding motif. In most cases the helical hooks comprise the most C-terminal region of the protein in question, while the repressors from TP901-1 and *Salmonella* are different, by having additional C-terminal regions. Even though both the TP901-1 CI/MOR system and the *Salmonella* phage SPC32H Rep/Ant systems both control lysogeny, the mechanisms are quite different. For example, the C-terminal regions are quite different, both structurally but also functionally. SPC32H Rep has a CDD C-terminal anti-repressor-binding domain (CAD, 161-198) [41] while so far all suggestions for the MOR binding region on CI are on its N-terminal domain[10,45]. Furthermore the Rep/Ant system is RecA/LexA mediated, while the TP901-1 switch is RecA/LexA independent. It seems that bacterial and phage regulators can easily build up new and diverse regulatory systems by mixing and matching of small domains.

Using the previously solved structure of the NTD:$O_L$half-site complex[10] and the newly determined CTD_1 structure together with SAXS data on O_L, CIΔ58 and CIΔ58:O_L we produce a model of the solution complex of this truncation construct with DNA. We show that CIΔ58 binds to adjacent half-sites on the palindromic O_L, with a decrease in the average distance between the NTDs. Thus we show unequivocally that TP901-1 CI binds canonically to DNA, unlike for example suggested for the *Salmonella* phage Rep, where the current hypothesis is that two monomers from different dimers in a tetramer bind to adjacent DNA half-sites[41].

In the SAXS reconstructed shape, the central dimerization region appears elongated, however a model created based on the crystal structure of CTD_1 and NTD:DNA complexes gives a good fit to the experimental data, unlike modeling of a coiled coil region. Previously, we carried out titration experiment using NMR with CIΔ58 with O_L[9]. Here we did not observe chemical shift changes in the CTD_1 region upon DNA binding indicating no major structural changes. CD spectroscopy led to detection of some changes on complexation, which could stem from small structural changes in the DNA or, when interpreted in terms of protein secondary structure, as a slight increase in helical content. In our previous NMR work we also found that the linker remains flexible when CIΔ58 is bound to DNA, also in agreement with CIΔ58:O_L SAXS data when inspecting the Kratky plot, which even in the bound state is indicative of a rather flexible protein,

though the dynamics of the NTDs are decreased. The flexible linker can allow for movement of the dimerization region away from the DNA binding region, which could be the reason for the *ab initio* model of the complex being more extended than the uncomplexed CIΔ58. Taking information from all biophysical techniques together, we conclude that upon complexation there are no major structural changes in the individual domains in CIΔ58, but there are changes in their respective orientations and subtle changes in flexibility and/or secondary structure propensity.

Funding and acknowledgements

This work was supported by the Danish Council for Independent Research (Grant 4002-00107) SAXS data were collected at BM29 at ESRF (Grenoble, France). Crystallographic data were collected at the ID30B and ID23-2 beamlines at ESRF (Grenoble, France) and the P11 beamline of DESY (Hamburg, Germany). Margit Pedersen is thanked for providing initial constructs. Mogens Kilstrup and Karin Hammer are acknowledged for fruitful discussions throughout the project. Travel to synchrotrons was supported by the DANSCATT program, funded by the Danish Ministry for Higher Education and Science (Grant 7055-00001B) and the European Community's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement N°283570). LLL, PWT, KKR and AV are members of ISBUC, Integrative Structural Biology at the University of Copenhagen (www.isbuc.ku.dk).

References

1 Twort FW (1915) AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES. *The Lancet* **186**, 1241–1243.

2 Duckworth DH (1976) Who discovered bacteriophage? *Bacteriol Rev* **40**, 793–802.

3 Oppenheim AB, Kobiler O, Stavans J, Court DL & Adhya S (2005) Switches in bacteriophage lambda development. *Annu. Rev. Genet.* **39**, 409–429.

4 Ptashne M (2004) *A Genetic Switch: Phage Lambda Revisited* CSHL Press.

5 Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, Melamed S, Leavitt A, Savidor A, Albeck S, Amitai G & Sorek R (2017) Communication between viruses guides lysis–lysogeny decisions. *Nature* **541**, 488–493.

6 Shearwin KE & Egan JB (2000) Establishment of lysogeny in bacteriophage 186. DNA binding and transcriptional activation by the CII protein. *J. Biol. Chem.* **275**, 29113–29122.

7 Pedersen M & Hammer K (2008) The role of MOR and the CI operator sites on the genetic switch of the temperate bacteriophage TP901-1. *J. Mol. Biol.* **384**, 577–589.

8 Nakanishi H, Pedersen M, Alsing AK & Sneppen K (2009) Modeling of the genetic switch of bacteriophage TP901-1: A heteromer of CI and MOR ensures robust bistability. *J. Mol. Biol.* **394**, 15–28.

9 Rasmussen KK, Frandsen KEH, Boeri Erba E, Pedersen M, Varming AK, Hammer K, Kilstrup M, Thulstrup PW, Blackledge M, Jensen MR & Lo Leggio L (2016) Structural and dynamics studies of a truncated variant of CI repressor from bacteriophage TP901-1. *Scientific Reports* **6**, 29574.

10 Frandsen KH, Rasmussen KK, Jensen MR, Hammer K, Pedersen M, Poulsen J-CN, Arleth L & Lo Leggio L (2013) Binding of the N-terminal domain of the lactococcal bacteriophage TP901-1 CI repressor to its target DNA: a crystallography, small angle scattering and NMR study. *Biochemistry*, 130902093647008.

11 Pedersen M, Lo Leggio L, Grossmann JG, Larsen S & Hammer K (2008) Identification of quaternary structure and functional domains of the CI repressor from bacteriophage TP901-1. *J. Mol. Biol.* **376**, 983–996.

12 Kelley LA & Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363–371.

13 McGeehan JE, Streeter SD, Thresh S-J, Taylor JEN, Shevtsov MB & Kneale GG (2011) Structural Analysis of a Novel Class of R–M Controller Proteins: C.Csp231I from Citrobacter sp. RFL231. *Journal of Molecular Biology* **409**, 177–188.

14 Kabsch W (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132.

15 Bibby J, Keegan RM, Mayans O, Winn MD & Rigden DJ (2012) AMPLE: A cluster-and-truncate approach to solve the crystal structures of small proteins using rapidly computed ab initio models. *Acta Crystallographica Section D: Biological Crystallography* **68**, 1622–1631.

16 Keegan RM, Bibby J, Thomas J, Xu D, Zhang Y, Mayans O, Winn MD & Rigden DJ (2015) Exploring the speed and performance of molecular replacement with AMPLE using QUARK ab initio protein models. *Acta Crystallographica Section D: Biological Crystallography* **71**, 338–343.

17 Xu D & Zhang Y (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins: Structure, Function and Bioinformatics* **80**, 1715–1735.
18 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658–674.
19 Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung L-W, Read RJ & Adams PD (2008) Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. *Acta Crystallographica Section D Biological Crystallography* **64**, 61–69.
20 Emsley P & Cowtan K (2004) Coot : model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography*, 2126–2133.
21 Emsley P & Lohkamp B (2010) Features and development of Coot research. *Acta Crystallographica Section D Biological Crystallography*, 486–501.
22 Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) *MolProbity* : all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography* **66**, 12–21.
23 Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
24 Laskowski RA (2001) PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.* **29**, 221–222.
25 Holm L & Rosenstrï¿½m P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Research* **38**, W545–W549.
26 Pedersen M, Lo Leggio L, Grossmann JG, Larsen S & Hammer K (2008) Identification of quaternary structure and functional domains of the CI repressor from bacteriophage TP901-1. *J. Mol. Biol.* **376**, 983–996.
27 Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I & Stockinger H (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research* **40**, W597–W603.
28 Konarev PV, Volkov VV, Sokolova AV, Koch MHJ & Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **36**, 1277–1282.
29 Guinier A & Fournet G (1955) *Small-angle scattering of X-rays* Wiley.
30 Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **25**, 495–503.
31 Porod G (1982) *In small angle x-ray scattering / edited by O. Glatter and O. Kratky* Academic Press, New York.
32 Svergun DI, Petoukhov MV & Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* **80**, 2946–2953.
33 Kozin MB & Svergun DI (2001) Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography* **34**, 33–41.
34 Volkov VV & Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **36**, 860–864.
35 Svergun DI (1999) Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophysical Journal* **76**, 2879–2886.

36 Lewis RJ, Brannigan JA, Offen WA, Smith I & Wilkinson AJ (1998) An evolutionary link between sporulation and prophage induction in the structure of a repressor:anti-repressor complex. *Journal of Molecular Biology* **283**, 907–912.
37 Colledge VL, Fogg MJ, Levdikov VM, Leech A, Dodson EJ & Wilkinson AJ (2011) Structure and Organisation of SinR, the Master Regulator of Biofilm Formation in Bacillus subtilis. *J Mol Biol* **411**, 597–613.
38 Rosenberg OS, Dovey C, Tempesta M, Robbins RA, Finer-Moore JS, Stroud RM & Cox JS (2011) EspR, a key regulator of Mycobacterium tuberculosis virulence, adopts a unique dimeric structure among helix-turn-helix proteins. *Proceedings of the National Academy of Sciences* **108**, 13450–13455.
39 Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D & Bryant SH (2013) CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Research* **41**, D348–D352.
40 Wang C, Ye F, Kumar V, Gao Y-G & Zhang L-H (2014) BswR controls bacterial motility and biofilm formation in Pseudomonas aeruginosa through modulation of the small RNA rsmZ. *Nucleic Acids Res.* **42**, 4563–4576.
41 Kim M, Kim HJ, Son SH, Yoon HJ, Lim Y, Lee JW, Seok Y-J, Jin KS, Yu YG, Kim SK, Ryu S & Lee HH (2016) Noncanonical DNA-binding mode of repressor and its disassembly by antirepressor. *Proceedings of the National Academy of Sciences* **113**, E2480–E2488.
42 Kim M & Ryu S (2013) Antirepression System Associated with the Life Cycle Switch in the Temperate Podoviridae Phage SPC32H. *Journal of Virology* **87**, 11775–11786.
43 Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **25**, 495–503.
44 Sreerama N & Woody RW (2000) Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set. *Analytical Biochemistry* **287**, 252–260.
45 Pedersen M, Ligowska M & Hammer K (2010) Characterization of the CI repressor protein encoded by the temperate lactococcal phage TP901-1. *J. Bacteriol.* **192**, 2102–2110.

Table 1 Crystallographic data and refinement statistics

Data and processing statistics	**CTD_1 – tetragonal**	**CTD_1 – orthorhombic**
Beamline	P11 (DESY)	P11 (DESY)
Wavelenght, λ [Å]	1.0332	1.0332
Space group	$P4_12_12$	$P2_12_12_1$
Unit cell dimensions		
(a, b, c) [Å]	42.95, 42.95, 102.50	43.84, 42.56, 101.16
(α, β, γ) [°]	90.0, 90.0, 90.0	90.0, 90.0, 90.0
No. of frames	3,600	3,600
Resolution range [Å]	42.9-1.81 (1.92-1.81)	40.23-1.50 (1.54-1.50)
Completeness [%]	100.0 (99.9)	99.9 (98.7)
R_{meas} [%]	10.0 (71.1)	18.3 (82.8)
I/σ(I)	26.21 (4.64)	15.22 (2.74))
$CC_{1/2}$ [%]	100.0 (90.7)	99.8 (74.2)

Observed reflections	224,742 (34,959)	372,711 (25,129)
Unique reflections	9,314 (1,457)	31,189 (2,226)
Redundancy	24.13 (23.99)	11.95 (11.29)
Refinement and validation statistics		
Resolution range [Å]		40.23-1.50 (1.55-1.50)
R_{work} [%]		14.55 (16.06)
R_{free} [%]		20.01 (20.64)
No. of protein atoms		1,819
No. of solvent atoms		228
RMSD of bond lengths [Å]		0.018
RMSD of bond angles [°]		1.630
Ramachandran statistics		
Favoured		99.49
Allowed		0.51
Outliers		0.00
Rotamer Outliers		0.00
Wilson B factor [$Å^2$]		10.4

Table 2. Analysis of experimental SAXS data of CIΔ58, O_L, CIΔ58:O_L and bovine serum albumin (BSA).

	R_g[a] (nm)	R_g[b] (nm)	V_p[c] (nm^3)	M_{wExp}[d] (kDa)	M_{wExp}[e] (kDa)	M_w[f] (kDa)
CIΔ58	2.82	2.93	46.83	26.63	26.90	30.088
O_L	1.91	1.98	15.66	-	-	11.614
CIΔ58:O_L	2.88	2.93	53.08	36.63	36.62	41.702
BSA	2.48	2.76	95.88	71.75	74.35	66.46

[a]Radius of gyration, R_g, estimated from the Guinier approximation.

[b]R_g estimated by AutoGNOM.

[c]The excluded hydrated volume (Porod volume), V_p.

[d]Molecular weight, M_w, estimated from $I(0)$ obtained by the Guinier approximation.

[e]M_w estimated from AutoGNOM.

[f]Theoretical M_w.

Figures

Figure 1: A) Structure of CI-CTD$_1$ showing the helical hook dimerization motif with the termini labeled. B) Structure of CI-CTD$_1$ showing the interacting residues. Residues forming salt bridges and hydrogen bonds are labeled and shown as sticks and the interactions are indicated by black lines. Chain A and B are colored orange and blue, respectively, in both panels.

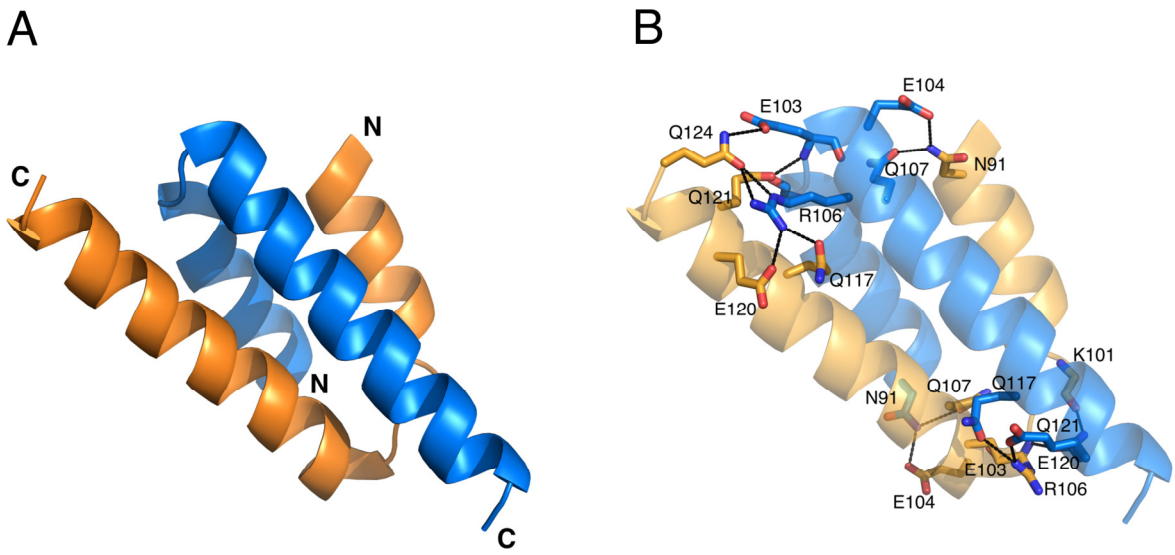


Figure 2: Isolation of CIΔ58:O_L complex. Purified CIΔ58 (asterisk, A) and freshly annealed O_L were mixed in a ratio (dimer to ds oligonucleotide) of 1:1.2 respectively. The mixture was incubated at room temperature for an hour, and loaded onto a Superdex75 Prepgrad. B) In circles (blue (A_{280}) and red (A_{260})) is the elution profile of the loaded mixture. C and D shows calculated A_{260}/A_{280} ratios calculated from the respective absorption spectra in A and B, respectively. Linear regression equation f(x) = -3.8875*K_{av}+11.6524 was used for M_w estimation. The equation was previously derived[9] using globular standards Conalbumin (75 kDa), Ovalbumin (43 kDa), Carbonic Anhydrase (29 kDa), Ribonuclease A (13.7 kDa) (Ge Health Care) and the well characterized globular domain CI-NTD (9.6 kDa). The K_{av}'s were calculated with V_o =109.65 mL, V_t = 320 mL, and elution volumes for each peak (V_e). From the K_{av} values masses were estimated K_{av}; CIΔ58 = 0.3152, K_{av}'s; Peak 1 = 0.2230, and Peak 2 = 0.3384. From the K_{av} values the following masses were estimated: M_w CIΔ58 =33.76 kDa and Mw Peak 1 = 48.31 kDa, and Peak 2 = 30.85 kDa.

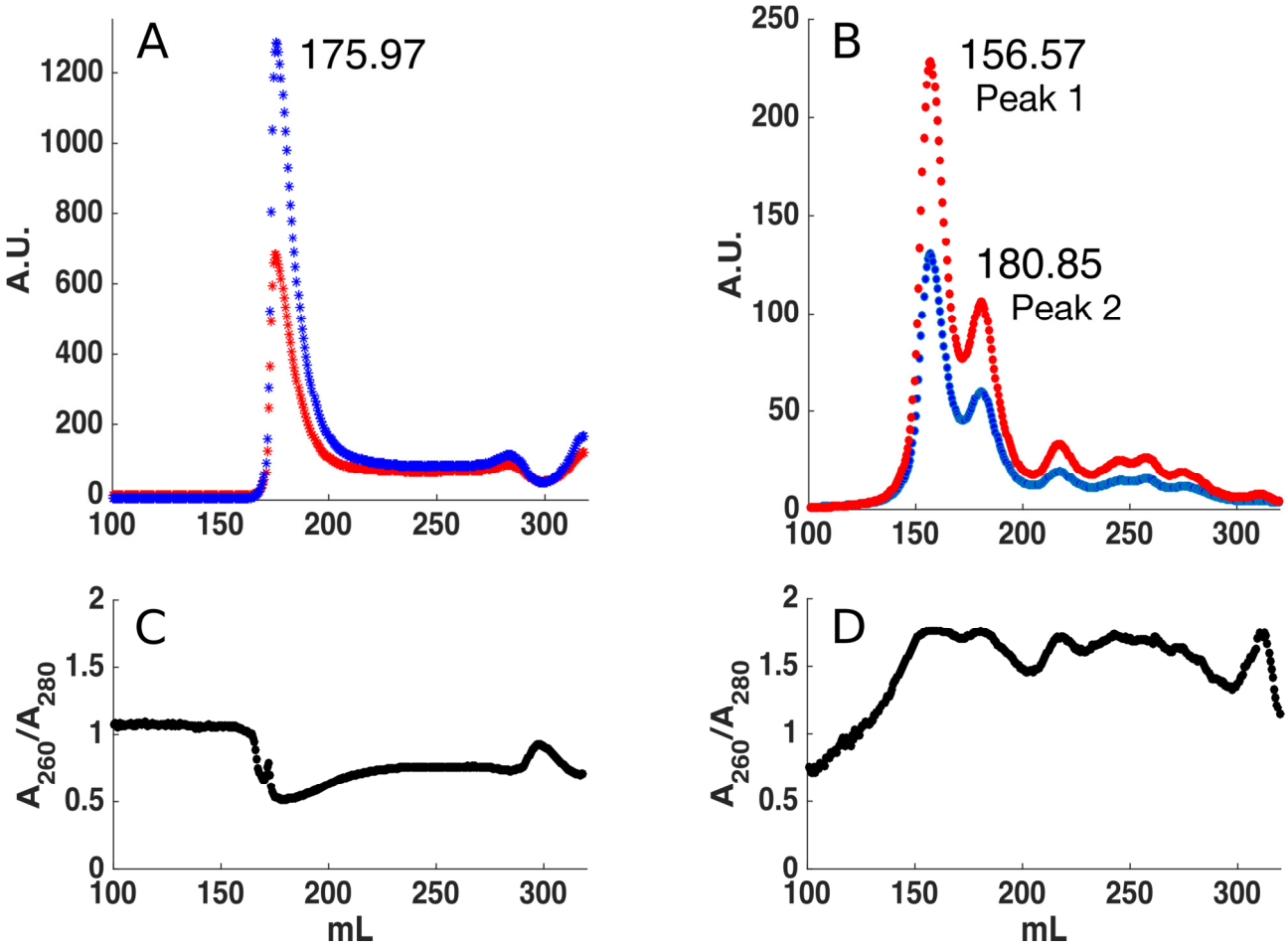


Figure 3: Analysis of SEC samples. To determine in which fraction the CIΔ58:O_L complex was eluded we ran a 6 % retardation gel (Thermofish) and stained it with Ethidiumbromide (upper gel) and InstantBlue® (lower gel). The single component annealed O_L (100 or 50 times diluted, 1 and 2, respectively) and purified CIΔ58 (4) for comparison. The purified CIΔ58 did not enter the gel due to it pI (theoretical pI = 8.00). Sample from the secondary SEC peak (180.85 mL, lane 6) consist mainly of O_L, and a small fraction of CIΔ58:O_L. Peak 1 at 156.57 mL (lane 7, 8 and 9 at 0.75 mg/mL, 1.3 mg/mL and 2.6 mg/mL, respectively) contains pure CIΔ58:O_L as no free O_L is observed.

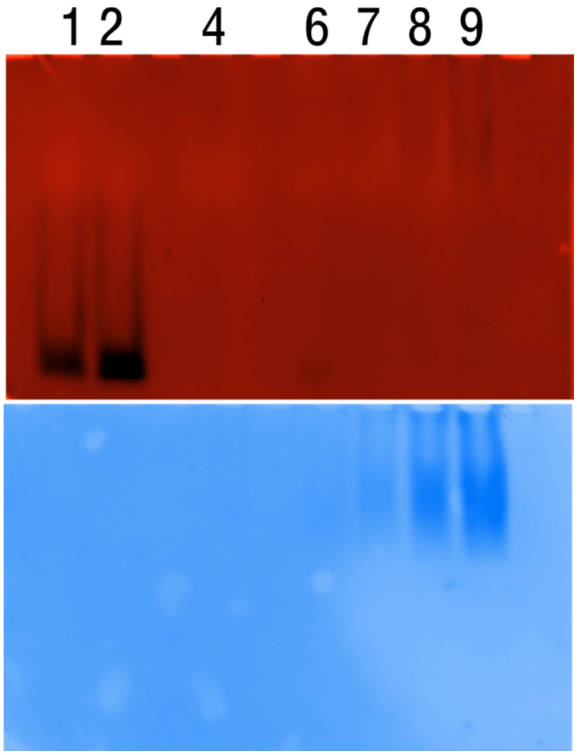


Figure 4: Experimental scattering curves of CIΔ58:O_L (A, red), CIΔ58 (A, blue) and O_L (A, yellow) were used to calculate the distance distribution (B, with same color code as for A). The O_L has maximum at ~1.9 nm (yellow, B), which corresponds to the diameter of the DNA helix. As found before, CIΔ58 has two maxima ~2 nm and ~4-5 nm corresponding to the cross-section and distances between the CTD_1 and NTDs and distances between the NTDs[9]. The distance plot of CIΔ58:O_L is clearly different from CIΔ58 alone. It has a single maximum at ~2.5 nm (the cross-section) but is also broader in general compared to CIΔ58. Comparing *ab initio* models clearly reveals differences between unbound (C, blue) and bound CIΔ58 (D, red). The most significant difference is the distance between NTDs from unbound with an average distance ~ 5.5 nm to ~3.0 nm in the bound form as determined from the center of masses, and is in good agreement with the distance between two major grooves of DNA (3.4 nm). The structural changes illustrate how flexibility enables CIΔ58 to 'grab' the DNA (grey) from opposing sides. Also significant is the density in the C-terminal part of the complex (CTD_1). Using high-resolution structures (pdb entries: 6FXA and 3ZHM) a structural model of CIΔ58 bound to O_L could be constructed (E).

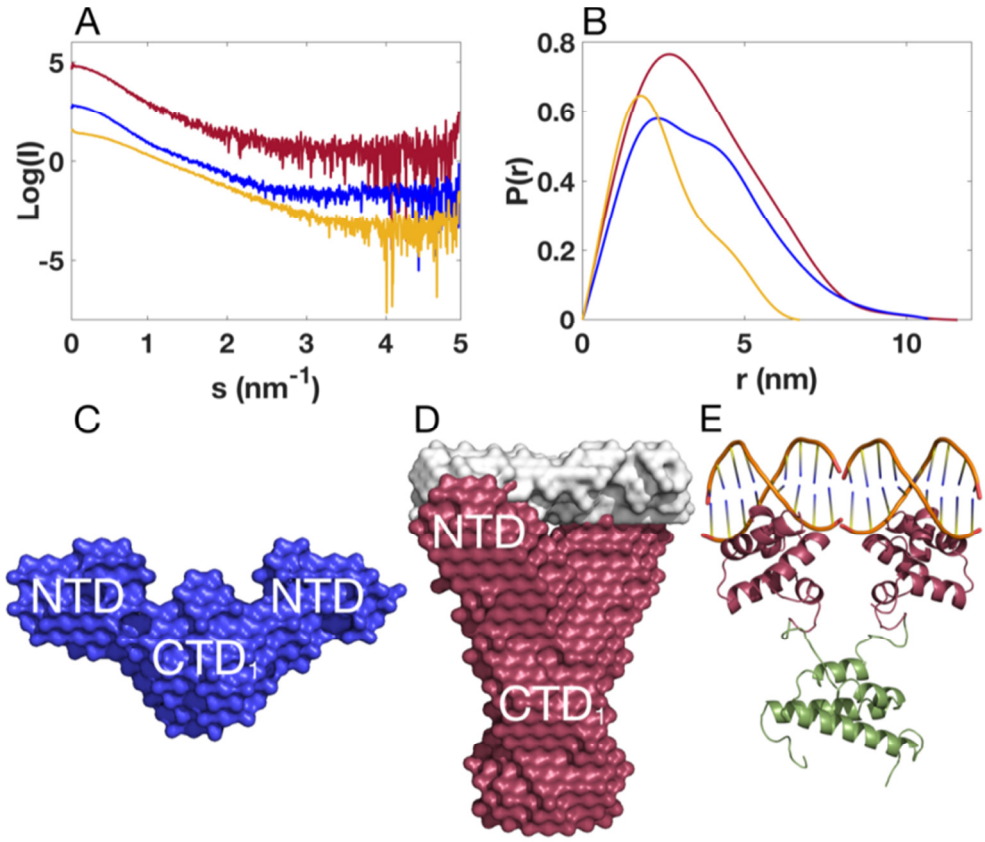
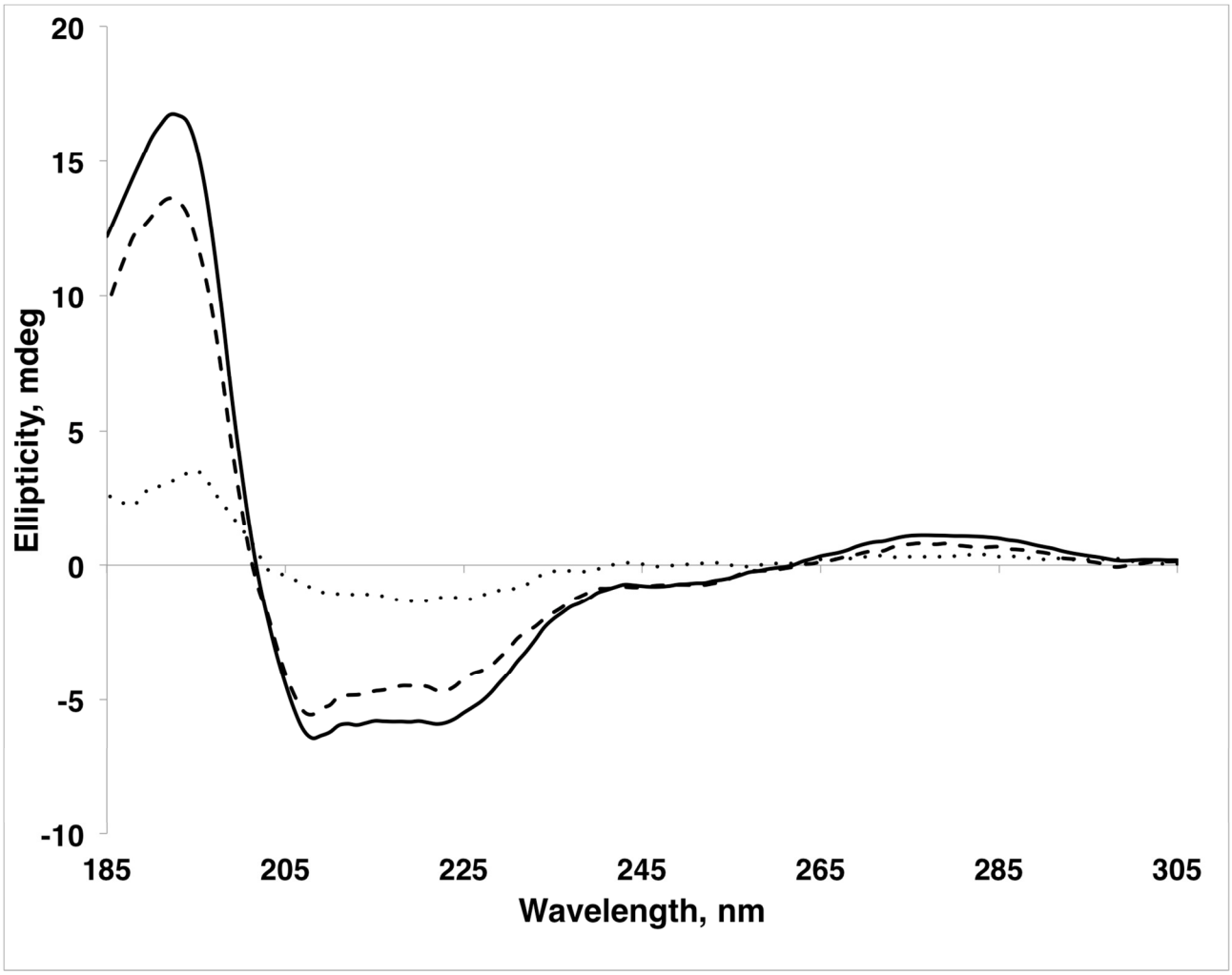


Figure 5: CD spectra of CIΔ58 with O_L measured in a 1 mm cell at room temperature in 10 mM Tris buffer with 100 mM NaF at pH 7.5. The solid line shows the experimental spectrum of CIΔ58 in complex with O_L, both at 19μM concentration. The dashed line shows the sum of the signals from corresponding spectra of CIΔ58 and O_L measured individually. The dotted line shows the difference between the complex and the sum. Component spectra are shown in Fig. S8.



Supplementary Information

Structural basis of the TP901-1 CI repressor dimerization and interaction with DNA

Kim Krighaar Rasmussen[1#], Anders K. Varming[1#], Simon N. Schmidt[1], Kristian E.H. Frandsen[1$], Peter W. Thulstrup[1], Malene Ringkjøbing Jensen[2], Leila Lo Leggio[1,*]

[1]Department of Chemistry, University of Copenhagen, Universitetsparken 5, Copenhagen, Denmark

[2]Univ. Grenoble Alpes, CNRS, CEA, IBS, F-38000 Grenoble, France

[$] Current address: INRA, UMR 1163 BBF (Biodiversité et Biotechnologie Fongiques), Polytech Marseille, Aix-Marseille Université, 163 Avenue de Luminy, 13288 Marseille Cedex 9, France

# equally contributing authors

[*] To whom correspondence should be addressed

Leila Lo Leggio

E-mail: leila@chem.ku.dk

Tel: +45 35 32 02 95

Running title: Characterization of CI-O_L repressor interaction

Keywords: phage repressor; protein-DNA complex; helical hook motif; dimerization domain; transcription factor

Superpositioning					PDBsum					
Protein	Residues	Keyword	RMSD (dimer)	Aligned res	Protein	Interface residues	Interface area, Å^2	Salt bridges	Hydrogen bonds	Non-bonded contacts
TP901-1 CTD1	90-124	Dimerization domain of CI repressor	-	-	**TP901-1 CTD1**	25:25	1324:1326	2	15	154
3IVP	82-114	DNA-binding protein from *Clostridium difficile* 630	1.4	64	**3IVP**	22:19	1125:1143	11	9	124
3LIS	69-94	Restriction-modification controller protein C.Csp231I	1.9	48	**3LIS**	10:10	739:738	-	-	32
3OP9	78-110	Transcriptional regulator from *Listeria innocua*	3.9	64	**3OP9**	13:13	804:804	2	2	46
1B0N	SinR 74-106 / SinI 9-39	SinR/SinI	4.4	56	**1B0N**	20:21	1333:1255	3	6	125
3QYX	98-133	EspR	2	64	**3QYX**	18:17	1165:1206	2	2	54
2YAL	71-109	SinR CTD dimer	2.9	64	**2YAL**	23:23	1443:1387	5	6	120
1OV9	2-46	Cholerae gene expression modulator	2.2	38	**10V9**	23:27	1616:1570	3	8	102
4O8B	68-101	bswR	2.2	64	**4O8B**	21:21	1068:1069	2	2	98
5D4Z	169-196	Salmonella repressor	2.2	48	**5D4Z**	9:10	578:605	-	-	40

Table S1. Structural comparison of helical hooks motives. The table is divided into a superpositioning and PDBsum[1] section. The superpositioning was conducted with the TP901-1 CTD_1 dimer as target using the cealign command in Pymol for all except for 1OV9, which was aligned with SSM Superpose in Coot. The PDBsum results show the interactions between the two chains that stabilize the respective dimer in each motif. Salt bridges are also included in the number of hydrogen bonds if they satisfy the criteria of PDBsum.

CIΔ58 without DNA		Helix		Turns	Others			
Reference set	Range, nm	Helix1	Helix2	Turns	Strand1	Strand2	Unordered	Total
3	185-260	57.00%	29.00%	4.00%	1.00%	2.00%	7.00%	100.00%
4	190-240	52.00%	23.00%	6.00%	3.00%	2.00%	13.00%	99.00%
6	185-240	54.00%	29.00%	7.00%	2.00%	2.00%	5.00%	99.00%
7	190-240	52.00%	29.00%	4.00%	3.00%	1.00%	10.00%	99.00%
Average total			**81.25%**	**5.25%**			**12.75%**	

CIΔ58 with DNA		Helix		Turns	Others			
Reference set	Range, nm	Helix1	Helix2	Turns	Strand1	Strand2	Unordered	Total
3	185-260	62.00%	28.00%	3.00%	2.00%	1.00%	5.00%	101.00%
4	190-240	60.00%	22.00%	7.00%	2.00%	1.00%	8.00%	100.00%
6	185-240	62.00%	28.00%	3.00%	1.00%	1.00%	4.00%	99.00%
7	190-240	56.00%	30.00%	4.00%	3.00%	1.00%	6.00%	100.00%
Average total			**87.00%**	**4.25%**			**8.75%**	

Table S2. CD analysis with DichroWeb. Shematic overview of the DichroWeb[2] analysis of the CD spectra of CIΔ58 alone and with DNA. The results have been averaged and grouped together in helix, turns and others.

Figure S1

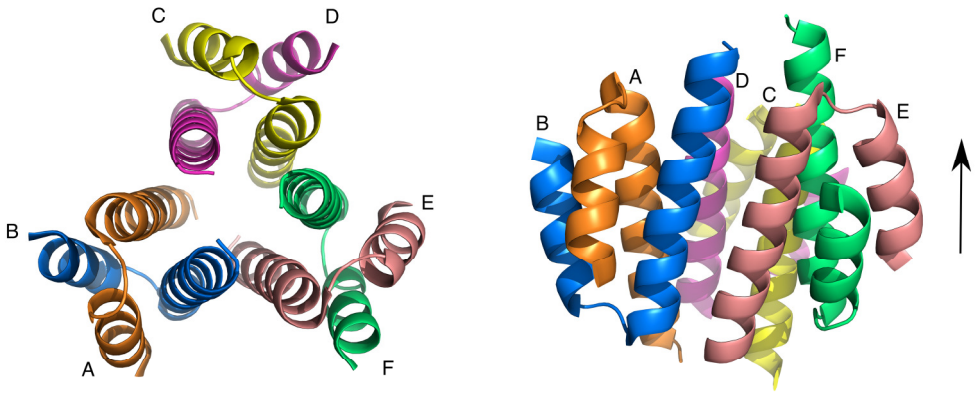


Figure S1. Cartoon representation of the arrangement of the trimer of dimers in the asymmetric unit. Left: Top view of the trimer of dimers with each chain indicated by its respective letter. Right: Side view of the trimer of dimers with same labelling as top. The arrow indicates the upward shift of the E/F dimer, which makes the trimer asymmetric.

Figure S2

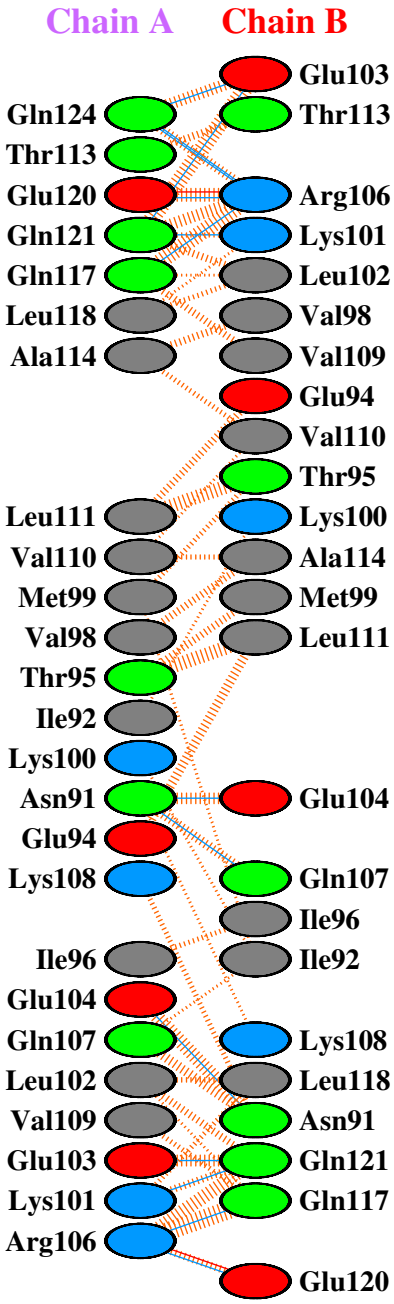


Figure S2. Interactions across the interfaces of chains A and B. Red lines indicate salt bridges, blue hydrogen bonds and orange dashed lines indicate non-bonded contacts. For non-bonded contacts, the width of the dashed line is proportional to the number of atomic contacts. The residues are colored blue for positive, red for negative, green for neutral and grey for aliphatic. Generated with PDBsum [1].

Figure S3

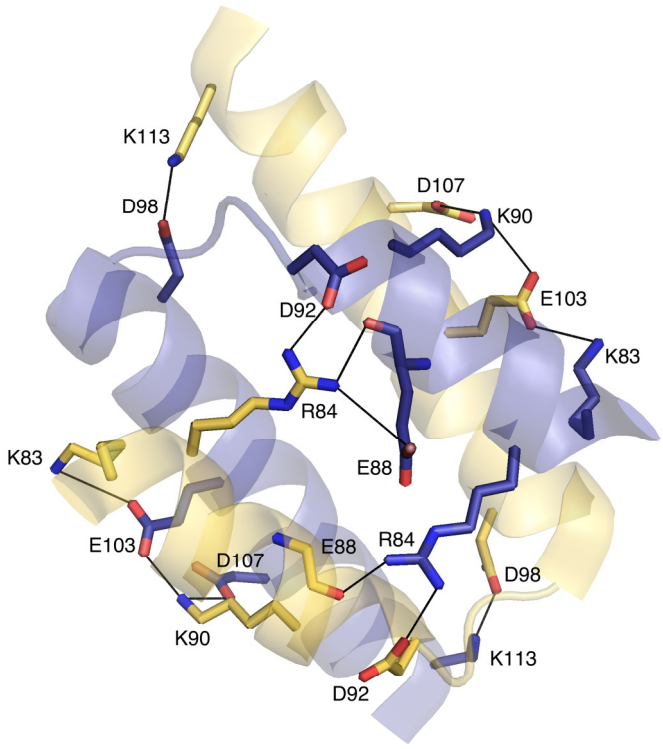


Figure S3. Structure of PDB:3IVP showing the interacting residues. Residues forming salt bridges and hydrogen bonds are labeled and shown as sticks and the interactions are indicated by black lines. Chain A and B are colored purple and yellow, respectively.

Figure S4

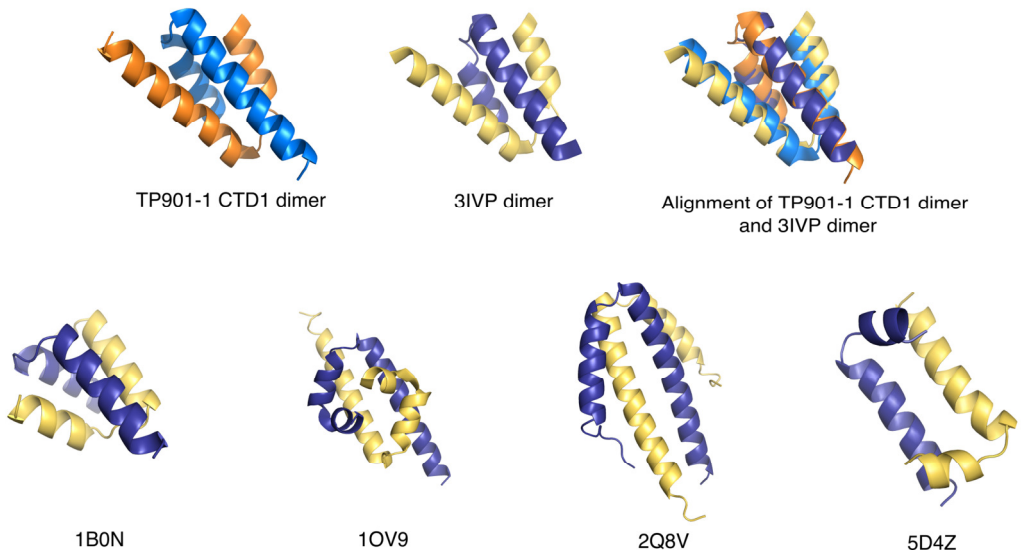


Figure S4. Illustration of different helical hooks structures mentioned in the main text. The structures are shown in the same orientation as CI-CTD$_1$

Figure S5

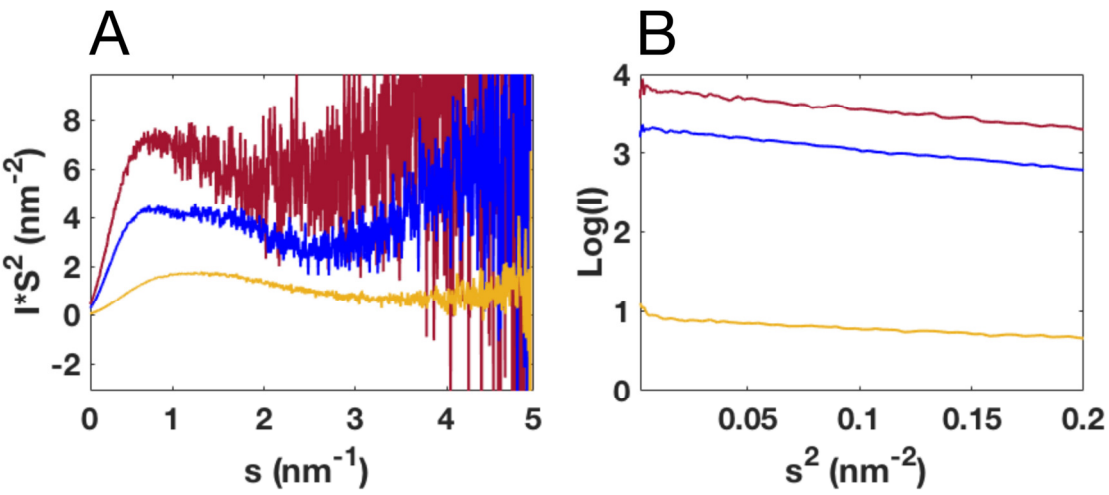


Figure S5. Analysis of small angle X-ray scattering data. The Kratky plot (A) shows that both CIΔ58:O_L (red) and CIΔ58 (blue) are flexible. However, the dynamics of the multi domain scattering particle are changed when CIΔ58 is bound to O_L, as seen from the loss of the second maxima in CIΔ58 (blue, A). As expected, O_L is a rigid scattering particle (yellow, A). From background subtracted scattering profiles (Figure 4 D) of CIΔ58:O_L (red), CIΔ58 (blue) and O_L (yellow), the Guinier regions of all scattering particles were determined with high fidelity, and as seen from the plot (B) a long linear region was obtained all cases.

Figure S6

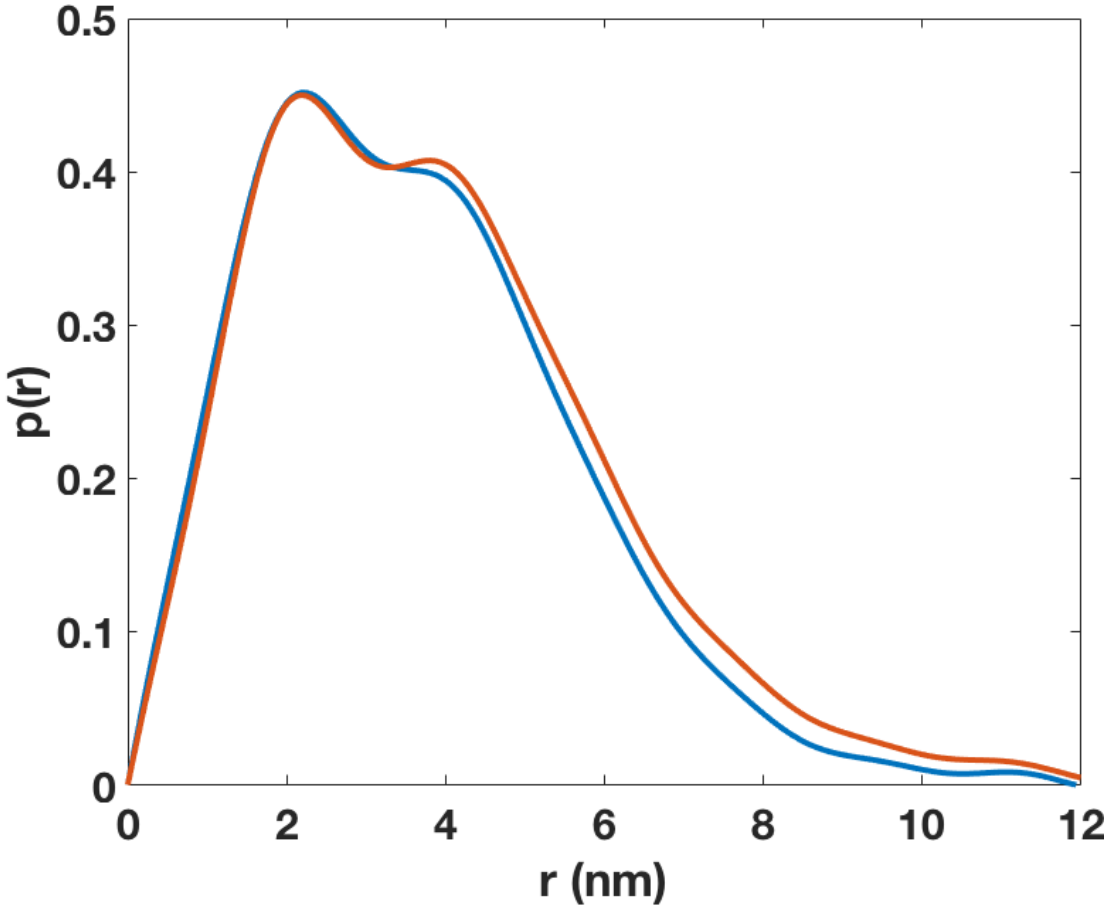
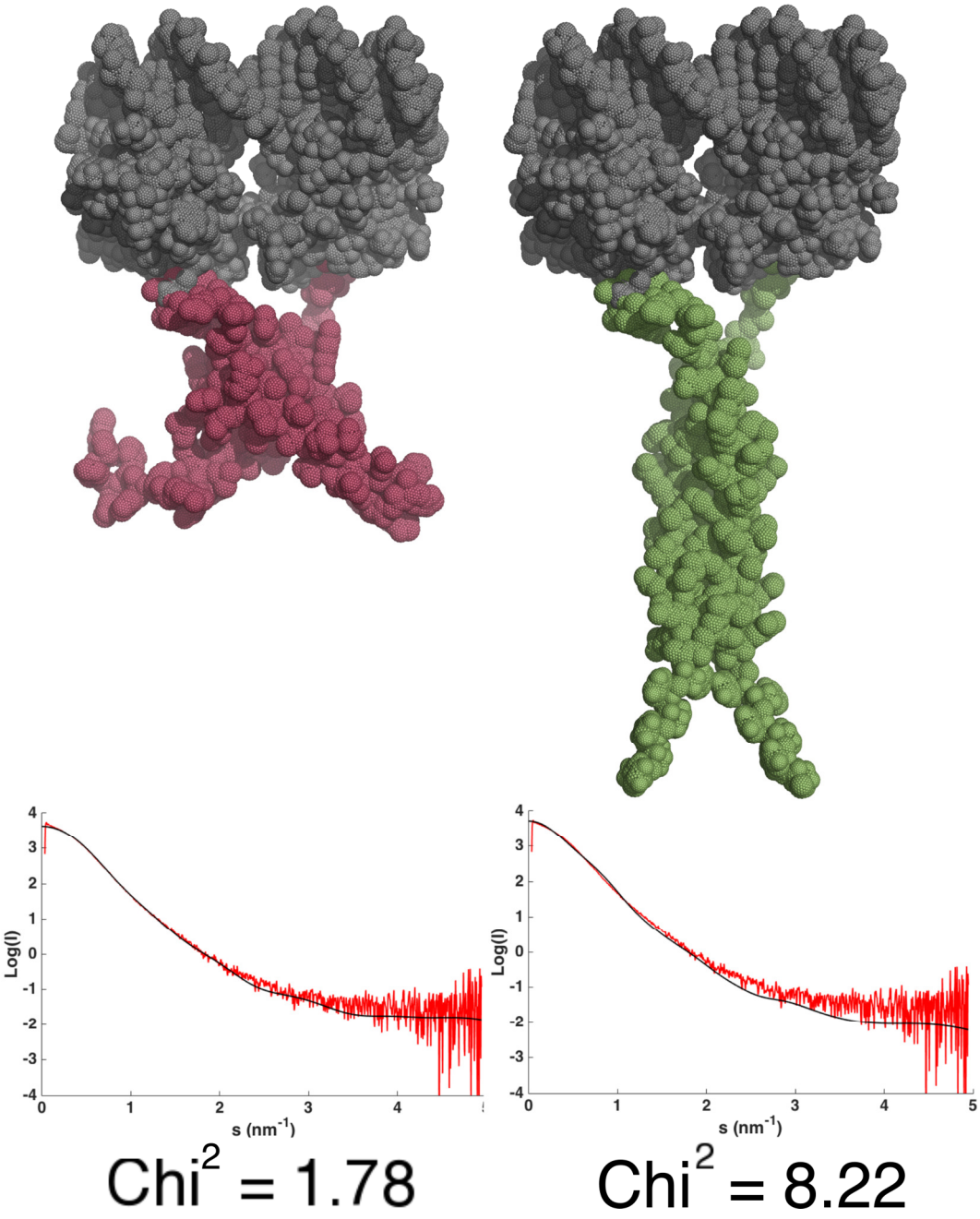


Figure S6. Comparing SAXS data recorded at +4 °C with 0.1 M NaCl (blue) or +25 °C with 1 M NaCl (red)[3]. P(r) curves of scattering curves recorded at +4 or +25 °C were calculated using GNOM[4], and plotted together to compare. When comparing them it is visible that both have similar maxima. The global maximum (r ~2 nm), which is the cross section of the scattering particle identical, whereas the local maximum (r ~4-5 nm) is less intense for the +4 °C data set compared to the +25 °C, suggesting that the temperature affects the dynamics in the scattering particle, and thus being less extended at +4 °C.

Figure S7



1.51 mg/mL	Chi^2 = 1.777	Chi^2 = 8.215
0.33 mg/mL	Chi^2 = 1.339	Chi^2 =1.840

Figure S7. Determining the conformation of the dimerization region. Two models consisting of either a helical-hooks (red) or a coiled coil (green) dimerization domain were used to calculate the theoretical scattering curves (black line) and fitted against experimental scattering data (red) by CRYSOL[5].

Figure S8

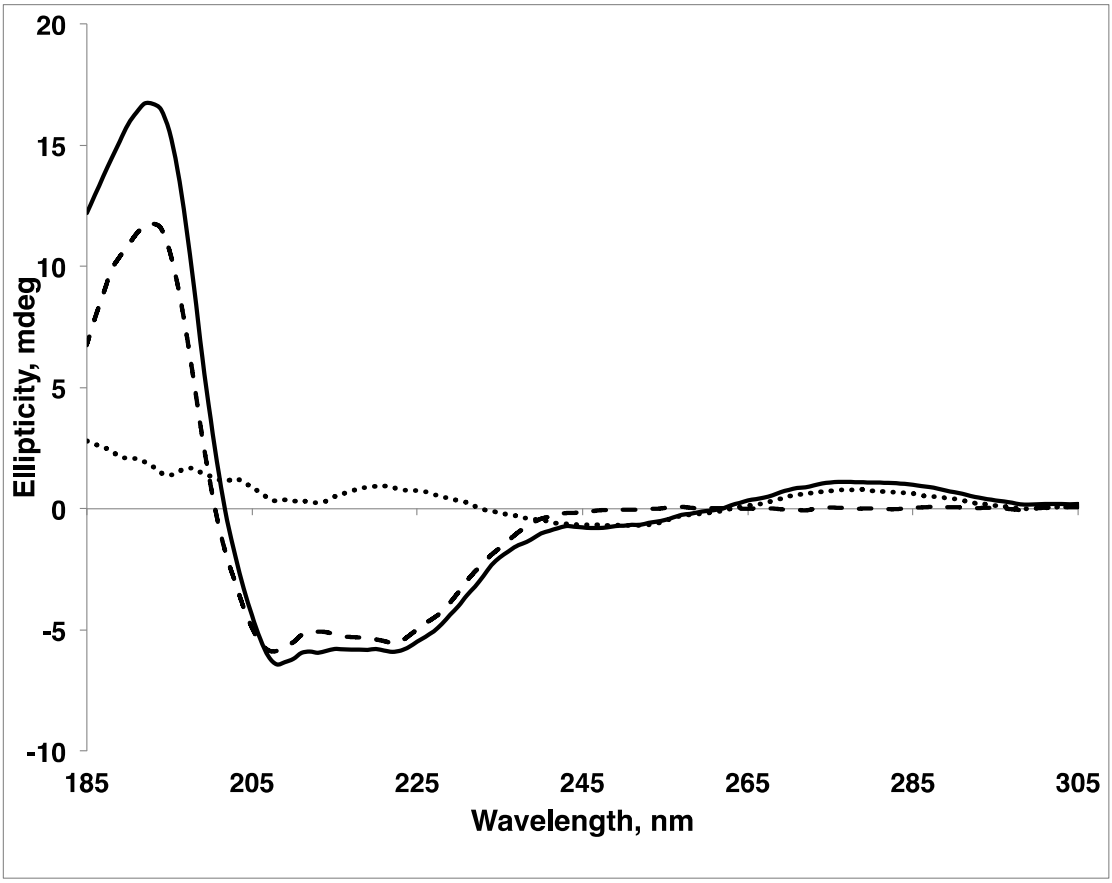


Figure S8. Circular dichroism spectra of CIΔ58 and O_L measured in a 1 mm cell at room temperature in 10 mM tris buffer with 100 mM NaF at pH 7.5. The solid line shows the experimental spectrum of CIΔ58 in complex with O_L, while the dashed and dotted lines show the individual experimental spectra of CIΔ58 and O_L, respectively. All spectra were measured at a 19 μM concentration for both CIΔ58 and O_L with subsequent buffer substraction.

References

1 Laskowski RA (2001) PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.* **29**, 221–222.

2 Whitmore L & Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research* **32**, W668–W673.

3 Rasmussen KK, Frandsen KEH, Boeri Erba E, Pedersen M, Varming AK, Hammer K, Kilstrup M, Thulstrup PW, Blackledge M, Jensen MR & Lo Leggio L (2016) Structural and dynamics studies of a truncated variant of CI repressor from bacteriophage TP901-1. *Scientific Reports* **6**, 29574.

4 Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **25**, 495–503.

5 Svergun D, Barberato C & Koch MHJ (1995) CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* **28**, 768–773.